Regression methods in pricing American and Bermudan options using consumption processes

Belomestny, Denis; Spokoiny, Vladimir; Milstein, Grigori

QUANTITATIVE FINANCE

Regression methods in pricing American and Bermudan options using consumption processes

Journal:	*Quantitative Finance*
Manuscript ID:	RQUF-2006-0209.R2
Manuscript Category:	Research Paper
Date Submitted by the Author:	12-Dec-2007
Complete List of Authors:	Belomestny, Denis; WIAS Spokoiny, Vladimir; WIAS Milstein, Grigori; Ural State University, Mathematics
Keywords:	American Options, Consumption, Monte Carlo Methods, Statistical Methods, LIBOR Market Models
JEL Code:	G12 - Asset Pricing < G1 - General Financial Markets < G - Financial Economics

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

rquf-2006-0209_revised_II.tex
RQUF-2006-0209.R2 Final Files.zip

Regression methods in pricing American and Bermudan options using consumption processes

Belomestny, Denis *
Weierstrass-Institute
Mohrenstr. 39, 10117 Berlin, Germany
belomest@wias-berlin.de

Milstein, Grigori N.†
Ural State University
Lenin Str. 51, 620083 Ekaterinburg, Russia
Grigori.Milstein@usu.ru

Spokoiny, Vladimir
Weierstrass-Institute
Mohrenstr. 39, 10117 Berlin, Germany
spokoiny@wias-berlin.de

Abstract

Numerical algorithms for efficient pricing multidimensional discrete-time American and Bermudan options are constructed using regression methods and a new approach for computing upper bounds of the options' price. Using the sample space with payoffs at the optimal stopping times, we propose sequential estimates for continuation values, values of the consumption process, and stopping times on the sample paths. The approach allows constructing both lower and upper bounds for the price by Monte Carlo simulations. The algorithms are tested by pricing Bermudan max-calls and swaptions in the Libor market model.

Keywords: American and Bermudan options, Error bounds, Monte Carlo, Consumption process, Regression methods, Optimal stopping times

AMS 2000 Subject Classification: 60H30, 65C05, 91B28

*The partial support of DFG through SFB 649 is gratefully acknowledged.

†This work was finished while the author was a visitor at the Weierstrass-Institute für Angewandte Analysis und Stochastik (WIAS), Berlin, due to the financial supports from this institute and DFG (grant No. 436 RUS 17/137/05 and 436 RUS 17/24/07) which are gratefully acknowledged.

1 Introduction

Valuation of high-dimensional American and Bermudan options is one of the most difficult numerical problems in financial engineering. Besides its practical relevance, this problem is of great theoretical importance since pricing American-style options is an archetype for high-dimensional optimal stopping problems. Several approaches have recently been proposed for pricing such options using the Monte Carlo technique (see, e.g. [1]-[15], [17]-[21], [25, 26, 28] and references therein). In practice it is often an open question whether the obtained numerical result is sufficiently accurate. As a rule, any numerical procedure has errors of various types (e.g., discretization or Monte Carlo errors) and it is difficult to take all of them into account. That is why in a number of works (see, e.g. [3, 4, 6, 9, 17, 18, 20, 25, 26]) some procedures are proposed to produce upper bounds along with lower bounds for the option price. The knowledge of lower and upper bounds makes it possible to evaluate accuracy of the price estimates. In [3] we developed an approach for pricing American options applicable both in the case of discrete-time and continuous-time financial models. The approach is based on the equivalence of an American option and a European one with consumption process (the so-called Earlier Exercise Premium representation, see [22]). It allows us, in principle, to iteratively construct a sequence v^1, V^1, v^2, V^2, v^3, ..., where v^1, v^2, v^3, ..., is an increasing sequence of lower bounds and V^1, V^2, ..., is a decreasing sequence of upper bounds. Unfortunately, the construction of the above sequence of bounds requires very laborious calculations. Even finding V^2 is, as a rule, too expensive. In [4] we proposed to use an increasing sequence of lower bounds for constructing both upper bound and lower bound at the initial position (t_0, X_0). It is assumed that this sequence is not too expensive from the computational point of view. This can be achieved by using local lower bounds which take into account the behavior of the process during a small number of steps ahead. The method of [4] is suitable for getting rough estimates. However, to obtain more accurate results, one needs rather expensive calculations.

Let us consider a discrete-time financial model

$$(B_{t_i}, X_{t_i}) = (B_{t_i}, X^1_{t_i}, ..., X^d_{t_i}),\ i = 0, 1, ..., \mathcal{I},$$

where B_{t_i} is the price of a scalar riskless asset and $X_{t_i} = (X^1_{t_i}, ..., X^d_{t_i})$ is the price vector of risky assets. Along with the index t_i, we shall use the index i, writing (t_i, X_i) instead of (t_i, X_{t_i}). Let $f_i(x)$ be a payoff at time t_i provided that $X_{t_i} = X_i = x,\ x \in \mathbf{X} \subset \mathbf{R}^d$, where $\mathbf{X}$ is a state space (e.g., $\mathbf{X} = \mathbf{R}^d$, $\mathbf{X} = \mathbf{R}^d_+$).

We assume that the modelling is based on the filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_i)_{0\le i\le \mathcal{I}}, P)$, where the probability measure P is the risk-neutral pricing measure for the problem under consideration, and X_i is a Markov chain with respect to the filtration $(\mathcal{F}_i)_{0\le i\le \mathcal{I}}$.

With respect to the probability measure P, the discounted process X_i/B_i is a mar-

tingale and the price $u_i(X_i)$ of the American option is given by

$$u_i(x) = \sup_{\tau \in \mathcal{T}_{i,\mathcal{I}}} B_i E \left(\frac{f_\tau(X_\tau^{t_i,x})}{B_\tau} \right). \tag{1.1}$$

In (1.1) $X_{t_j}^{t_i,x}$ is the value of the Markov chain at instant $t_j \geq t_i$ starting at t_i from x and $\mathcal{T}_{i,\mathcal{I}}$ is the set of stopping times τ taking values in $\{i, i+1, ..., \mathcal{I}\}$.

The value process u_i (Snell envelope) can be determined by induction as follows:

$$\begin{aligned} u_{\mathcal{I}}(x) &= f_{\mathcal{I}}(x), \\ u_i(x) &= \max \left\{ f_i(x), B_i E \left(\frac{u_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x \right) \right\}, \ i = \mathcal{I}-1, ..., 0. \end{aligned} \tag{1.2}$$

We see that theoretically the problem of evaluating $u_0(X_0)$, the price of the discrete-time American option at the initial position (t_0, X_0), can be easily solved using dynamic programming algorithm (1.2). However, if X is high dimensional and $\mathcal{I}$ is large, this algorithm is not practical.

In order to use regression methods for sequential evaluation of u_i, one can consider (see, e.g., [28] and [15]) the $(d+1)$-dimensional sample

$$\left({}_mX_i, \frac{B_i}{B_{i+1}} u_{i+1}({}_mX_{i+1}) \right), \ m = 1, ..., M, \ i = 0, ..., \mathcal{I}-1, \tag{1.3}$$

from $\left(X_i, \frac{B_i}{B_{i+1}} u_{i+1}(X_{i+1}) \right)$, where $(t_i, \ {}_mX_i)$ are M independent trajectories all starting from the point (t_0, X_0). The use of the procedure (1.2) and sample (1.3) for sequential evaluating $u_i(X_i)$ together with modern methods of multidimensional approximation (see e.g., [13], [29] and references therein) can give effective algorithms for pricing American and Bermudan options (see e.g. [5], [19]).

The samples with optimal stopping times $\tau^{t_i,x} = \tau^{i,x}$ were first introduced in [23] (see also [12] and [15]). Applying (1.3), one needs an estimate $\widehat{u}_{i+1}(X_{i+1})$ of $u_{i+1}(X_{i+1})$ while applying the samples with stopping times, we can employ an estimate $\widehat{\tau} = \widehat{\tau}^{t_{i+1}, \ X_{i+1}}$ of $\tau^{t_{i+1}, \ X_{i+1}}$. In the latter case the corresponding estimate for $f_\tau(X_\tau)$ is $f_{\widehat{\tau}}(X_{\widehat{\tau}})$ and the inequality

$$E \left(\frac{f_{\widehat{\tau}}(X_{\widehat{\tau}})}{B_{\widehat{\tau}}} \middle| X_i \right) \leq E \left(\frac{f_\tau(X_\tau)}{B_\tau} \middle| X_i \right)$$

obviously holds. This inequality opens the possibility to construct a lower bound for continuation value (lower continuation value). In turn, this allows us (see Section 2.3) to construct an upper bound for consumption process (upper consumption process). Thus, in contrast to other works using regression methods for pricing American and Bermudan options (see, however, [6]), we construct not only an estimate for continuation value but also an estimate for upper consumption process making it possible to find effectively lower and upper bounds for the price of the option.

In Section 2, we recall the approach (see [3], [4]) to pricing American and Bermudan options using consumption processes in the form suitable for our purposes. Furthermore,

we give a comparison with the dual approach (see [25], [17]) for the first time. In Section 3, we propose a number of algorithms for subsequent estimating optimal stopping times and continuation values using various regression methods. Special attention is paid to linear regression methods (see [23] and [12]). Section 4 gives formulas for the Monte Carlo construction of lower and upper bounds at the initial position (t_0, X_0). Section 5 is devoted to numerical experiments with Bermudan max-call and Bermudan swaption in a full factor Libor market model, which confirm efficiency of the proposed algorithms.

2 The approach based on consumption processes

To be self-contained, let us briefly recall the approach to pricing American and Bermudan options using consumption processes [3].

2.1 The continuation value, the continuation and exercise regions.

For the considered American option, let us introduce the continuation value

$$C_i(x) = B_i E\left(\frac{u_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right),\ i = 0, ..., \mathcal{I}-1;\ C_{\mathcal{I}}(x) = f_{\mathcal{I}}(x), \tag{2.1}$$

the continuation region $\mathcal{C}$ and the exercise (stopping) region $\mathcal{E}$:

$$\begin{aligned} \mathcal{C} &= \{(t_i, x) : f_i(x) < C_i(x)\}, \\ \mathcal{E} &= \{(t_i, x) : f_i(x) \geq C_i(x)\}. \end{aligned} \tag{2.2}$$

Clearly, $(t_{\mathcal{I}}, x) \in \mathcal{E}$ for any x.

Let $X_j^{i,x},\ j = i, i+1, ..., \mathcal{I}$, be the Markov chain starting at time step i from the point $x:\ X_i^{i,x} = x$, and ${}_mX_j^{i,x},\ m = 1, ..., M$, be independent trajectories of the Markov chain. The Monte Carlo estimator $\widehat{u}_i(x)$ for $u_i(x)$ (in the case when $\mathcal{E}$ is known) has the form

$$\widehat{u}_i(x) = \frac{1}{M}\sum_{m=1}^{M}\frac{B_i}{B_\tau} f({}_mX_\tau^{i,x}), \tag{2.3}$$

where τ is the first time at which $X_j^{i,x}$ enters $\mathcal{E}$ (of course, τ in (2.3) depends on i, x, and $m:\ \tau =_m \tau^{i,x}$). Thus, for estimating $u_i(x)$, it is sufficient to check at each time step t_j for $j = i, i+1, ..., \mathcal{I}$ whether the position $(t_j,\ {}_mX_j^{i,x})$ is in $\mathcal{E}$. Given a lower bound v, a simple sufficient condition for the continuation can be formulated. Introduce the following subset of the continuation region

$$\mathcal{C}_v = \left\{(t_k, x) : f_k(x) < B_k E\left(\frac{v_{k+1}(X_{k+1})}{B_{k+1}}|X_k = x\right)\right\}.$$

Since $\mathcal{C}_v \subset \mathcal{C}$, the condition $(t_j,\ {}_mX_j^{i,x}) \in \mathcal{C}_v$ is sufficient. It follows from (1.2) that $f_i(x)$ is a lower bound and if $v_i^1, ..., v_i^l$ are lower bounds then the function $v_i(x) = \max_{1\leq k\leq l} v_i^k(x)$ is a lower bound as well. Henceforth we consider lower bounds satisfying the inequality $v_i(x) \geq f_i(x)$.

2.2 Equivalence of American options to European ones with consumption processes involved

For $0 \le i \le \mathcal{I}-1$ the equation (1.2) can be rewritten in the form

$$u_i(x) = B_i E\left(\frac{u_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right) + \left[f_i(x) - B_i E\left(\frac{u_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right)\right]^+. \quad (2.4)$$

Introduce the functions

$$\gamma_i(x) = \left[f_i(x) - B_i E\left(\frac{u_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right)\right]^+, \ i = \mathcal{I}-1, ..., 0. \quad (2.5)$$

Due to (2.4), we get (see [3])

$$u_i(X_i) = B_i E\left(\frac{f_{\mathcal{I}}(X_{\mathcal{I}})}{B_{\mathcal{I}}}|\mathcal{F}_i\right) + B_i \sum_{k=1}^{\mathcal{I}-(i+1)} E\left(\frac{\gamma_{\mathcal{I}-k}(X_{\mathcal{I}-k})}{B_{\mathcal{I}-k}}|\mathcal{F}_i\right) \quad (2.6)$$
$$+\gamma_i(X_i), \ i = 0, ..., \mathcal{I}-1.$$

Putting $X_0 = x$ and assuming $B_0 = 1$, we obtain

$$u_0(x) = E\left(\frac{f_{\mathcal{I}}(X_{\mathcal{I}})}{B_{\mathcal{I}}}\right) + \gamma_0(x) + \sum_{i=1}^{\mathcal{I}-1} E\left(\frac{\gamma_i(X_i)}{B_i}\right). \quad (2.7)$$

Formula (2.7) gives the value of the European option with payoff function $f_{\mathcal{I}}(x)$ and consumption process γ_i defined by (2.5).

2.3 Upper and lower bounds using consumption processes.

Formula (2.7) cannot be used directly to value the discrete-time American option as the process $\gamma_i(x)$ is not known. In this section we describe how to construct lower and upper bounds for $u_i(x)$ (see [3] for more details).

Let $v_i(x)$ be a lower bound for the true option price $u_i(x)$. We introduce the functions (upper consumption processes):

$$\gamma_{i,v}(x) = \left[f_i(x) - B_i E\left(\frac{v_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right)\right]^+, \ i = 0, ..., \mathcal{I}-1. \quad (2.8)$$

Clearly,

$$\gamma_{i,v}(x) \ge \gamma_i(x).$$

Hence the price $V_i(x)$ of the European option with payoff function $f_{\mathcal{I}}(x)$ and upper consumption process $\gamma_{i,v}(x)$ is an upper bound: $V_i(x) \ge u_i(x)$.

Remark 2.1. Application of Jensen's inequality shows that if the expectation in (2.8) is estimated by the Monte Carlo method then the resulting estimate $\widehat{\gamma}_{i,v}(x)$ is upper biased, i.e., $E\widehat{\gamma}_{i,v}(x) \ge \gamma_{i,v}(x)$.

Indeed, let ${}_mX_{i+1}^{t_i,x}$, $m = 1, \dots, M$, be a set of one step trajectories all starting from (t_i, x). Defining

$$\widehat{\gamma}_{i,v}(x) := \left[f_i(x) - \frac{B_i}{M} \sum_{m=1}^{M} \frac{v_{i+1}({}_mX_{i+1}^{t_i,x})}{B_{i+1}} \right]^+, \tag{2.9}$$

we get by Jensen's inequality that $E\widehat{\gamma}_{i,v}(x) \geq \gamma_{i,v}(x)$.

Conversely, if $V_i(x)$ is an upper bound for the true option price $u_i(x)$ and

$$\gamma_{i,V}(x) = \left[f_i(x) - B_i E\left(\frac{V_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x \right) \right]^+, \; i = 0, ..., \mathcal{I} - 1, \tag{2.10}$$

then

$$\gamma_{i,V}(x) \leq \gamma_i(x),$$

and the price $v_i(x)$ of the European option with lower consumption process $\gamma_{i,V}(x)$ is a lower bound: $v_i(x) \leq u_i(x)$. The same reasoning that led to (2.9) implies now that the Monte Carlo estimate $\widehat{\gamma}_{i,V}(x)$ of $\gamma_{i,V}(x)$ is biased up. Therefore, for $\widehat{\gamma}_{i,V}(x)$ to be a lower bound, M must be large.

Thus, starting from a lower bound $v_i^1(x)$, one can construct the upper bound $V_i^1(x)$ as the European option with consumption process $\gamma_{i,v^1}(x)$ and so on. This procedure gives us the sequences $v_i^1(x) \leq v_i^2(x) \leq v_i^3(x) \leq ... \leq u_i(x)$, and $V_i^1(x) \geq V_i^2(x) \geq ... \geq u_i(x)$. All the bounds v^k and V^k can, in principle, be evaluated by Monte Carlo simulations. However, each further step of the procedure requires time-consuming calculations, and in practice it is possible to make only a few steps of this procedure. In this connection, much attention was given in [3] to variance reduction techniques and some constructive methods reducing statistical errors were proposed.

2.4 Comparison with the dual approach

Without loss of generality, we assume in this section that $B_i \equiv 1$. The dual approach, developed in [25] and [17], is based on the following observation. For any $0 \leq i \leq \mathcal{I}$ and any supermartingale $(S_j)_{i \leq j \leq \mathcal{I}}$ with $S_i = 0$, we have

$$\begin{aligned} u_i(X_i) = \sup_{\tau \in \mathcal{T}_{i,\mathcal{I}}} E\left(f_\tau(X_\tau)|\mathcal{F}_i\right) &\leq \sup_{\tau \in \mathcal{T}_{i,\mathcal{I}}} E\left(f_\tau(X_\tau) - S_\tau|\mathcal{F}_i\right) \\ &\leq E\left[\max_{i \leq j \leq \mathcal{I}} \left(f_j(X_j) - S_j\right) |\mathcal{F}_i \right]. \end{aligned} \tag{2.11}$$

Hence the right-hand side of (2.11) provides an upper bound for $u_i(X_i)$. It can be shown that the equality in (2.11) is attained at the martingale part of the Doob-Meyer decomposition of the price process u_i:

$$M_i = 0, \quad M_j = \sum_{l=i+1}^{j} \left(u_l(X_l) - E\left(u_l(X_l)|\mathcal{F}_{l-1}\right) \right), \quad i < j \leq \mathcal{I}.$$

The duality representation provides a simple way to estimate the Snell envelope from above, using a lower approximation process $\{v_i(X_i)\}$. Let M^v be the martingale

$$\begin{aligned} M_0^v &= 0; \qquad (2.12)\\ M_j^v &= M_{j-1}^v + v_j(X_j) - E\left(v_j(X_j)|\mathcal{F}_{j-1}\right)\\ &= \sum_{l=1}^{j} v_l(X_l) - \sum_{l=1}^{j} E\left(v_l(X_l)|\mathcal{F}_{l-1}\right), \quad 1 \le j \le \mathcal{I}. \end{aligned}$$

Then for any $0 \le i \le \mathcal{I}$ the process $\widetilde{M}_{ij} = M_j^v - M_i^v$, $j = i, \dots, \mathcal{I}$, is a martingale with $\widetilde{M}_{ii} = 0$ and according to (2.11)

$$V_i^D(X_i) := E\left[\max_{i \le j \le \mathcal{I}} \left(f_j(X_j) - \widetilde{M}_{ij}\right) |\mathcal{F}_i\right] \ge u_i(X_i).$$

In particular, for $i = 0$

$$\begin{aligned} V_0^D(X_0) &= v_0(X_0)\\ &+ E\left[\max_{0 \le j \le \mathcal{I}} \left(f_j(X_j) - v_j(X_j) + \sum_{l=0}^{j-1} \left(E\left(v_{l+1}(X_{l+1})|\mathcal{F}_l\right) - v_l(X_l)\right)\right)\right]. \qquad (2.13) \end{aligned}$$

The upper bound $V_0(X_0)$ obtained in Section 2.3 can be transformed to

$$\begin{aligned} V_0(X_0) &= E\left(f_{\mathcal{I}}(X_{\mathcal{I}})\right) + E\sum_{i=0}^{\mathcal{I}-1} \left[f_i(X_i) - E\left(v_{i+1}(X_{i+1})|\mathcal{F}_i\right)\right]^+\\ &= v_0(X_0) + E\sum_{i=0}^{\mathcal{I}-1} \left(\max\left\{f_i(X_i), E\left(v_{i+1}(X_{i+1})|\mathcal{F}_i\right)\right\} - v_i(X_i)\right), \qquad (2.14) \end{aligned}$$

where it is assumed that

$$f_i(X_i) \le v_i(X_i), \quad i = 0, \dots, \mathcal{I}-1, \quad v_{\mathcal{I}}(X_{\mathcal{I}}) = f_{\mathcal{I}}(X_{\mathcal{I}}).$$

It is interesting to compare V_0 and V_0^D starting from the same lower bound v_i. A comprehensive comparison of $V_0(X_0)$ and $V_0^D(X_0)$ seems to be difficult and we restrict ourselves to some examples. First, we construct examples where $V_0(X_0) \le V_0^D(X_0)$. Let us define

$$\tau := \min\left\{0 \le i \le \mathcal{I}-1 : f_i(X_i) \ge E\left(v_{i+1}|\mathcal{F}_i\right)\right\},$$

and $\tau = \mathcal{I}$ if $f_i(X_i) < E\left(v_{i+1}|\mathcal{F}_i\right)$ for all i. We see that if $\tau = \mathcal{I}$ or

$$f_i(X_i) \ge E\left(v_{i+1}(X_{i+1})|\mathcal{F}_i\right), \quad i \ge \tau,$$

with probability 1, then

$$\begin{aligned} V_0(X_0) &= v_0(X_0) + E\sum_{i=0}^{\tau-1} \left(E\left(v_{i+1}(X_{i+1})|\mathcal{F}_i\right) - v_i(X_i)\right)\\ &+ E(f_\tau(X_\tau) - v_\tau(X_\tau)) + E\sum_{j=\tau+1}^{\mathcal{I}-1} \left(f_j(X_j) - v_j(X_j)\right) \le V_0^D(X_0). \end{aligned}$$

The strict inequality $V_0 < V_0^D$ is achieved in the following simple example with $\mathcal{I} = 3$. Due to (2.11), the dual price at time 0 can be computed via the formula

$$\begin{aligned} V_0^D &= E\max\{f_0, f_1 - v_1 + Ev_1, \max\{f_2, E(u_3|\mathcal{F}_2)\} + Ev_1 + E(v_2|\mathcal{F}_1) - v_1 - v_2\} \\ &= E\max\{f_0, f_1 - v_1 + Ev_1, E(v_2|\mathcal{F}_1) + u_2 - v_2 - v_1 + Ev_1\} \\ &= E\max\{f_0, \max\{f_1, E(v_2|\mathcal{F}_1) + u_2 - v_2\} - v_1 + Ev_1\}, \end{aligned} \tag{2.15}$$

where we use the equality $u_2 = \max\{f_2, E(u_3|\mathcal{F}_2)\}$ and the dependence of quantities involved on the underlying process X_i is not shown explicitly for the sake of simplicity. Formula (2.14) gives

$$\begin{aligned} V_0 = E\max\{f_0, Ev_1\} &+ E(\max\{f_1, E(v_2|\mathcal{F}_1)\} - v_1) \\ &+ E(\max\{f_2, E(v_3|\mathcal{F}_2)\} - v_2). \end{aligned} \tag{2.16}$$

Let us take constant payoffs satisfying

$$f_0 < f_1 < f_2 < f_3, \quad f_1 + f_2 < f_0 + f_3.$$

Clearly, $u_i = f_3$, $i = 0, \ldots, 3$ and any lower bound v_i satisfies

$$f_0 \le v_0 \le f_3, \quad f_1 \le v_1 \le f_3, \quad f_2 \le v_2 \le f_3, \quad v_3 = f_3.$$

Formula (2.16) gives $V_0 = f_3$ and (2.15) implies

$$V_0^D = E\max\{f_0, E(v_2|\mathcal{F}_1) + f_3 - v_2 + Ev_1 - v_1\}.$$

Clearly,

$$V_0^D \ge E[E(v_2|\mathcal{F}_1) + f_3 - v_2 + Ev_1 - v_1] = f_3.$$

If v_1 and v_2 are such that the inequality

$$E(v_2|\mathcal{F}_1) + f_3 - v_2 + Ev_1 - v_1 \ge f_0$$

is fulfilled with probability 1, then $V_0^D = f_3$. However, if

$$E(v_2|\mathcal{F}_1) + f_3 - v_2 + Ev_1 - v_1 < f_0 \tag{2.17}$$

with positive probability, then

$$\max\{f_0, E(v_2|\mathcal{F}_1) + f_3 - v_2 + Ev_1 - v_1\} > E(v_2|\mathcal{F}_1) + f_3 - v_2 + Ev_1 - v_1$$

with the same probability and consequently $V_0^D > V_0$. The inequality (2.17) is achieved, for example, if Ev_1 is close to f_1, $E(v_2|\mathcal{F}_1)$ is close to f_2 and v_1 and v_2 are equal to f_3 with positive probability.

At the same time it is possible to construct examples when $V_0^D \le V_0$. Indeed, let us take $v_i(X_i) = f_i(X_i)$ for all $i = 0, \dots, \mathcal{I}-1$, then according to (2.11)

$$V_0^D = f_0 + E\left[\max_{0\le j\le \mathcal{I}} \sum_{l=0}^{j-1} \left(E\left(f_{l+1}|\mathcal{F}_l\right) - f_l\right)\right]$$

and due to (2.14)

$$V_0 = f_0 + \sum_{i=0}^{\mathcal{I}-1} \left(E\left(f_{i+1}|\mathcal{F}_i\right) - f_i\right)^+ \ge V_0^D.$$

However, the method based on the representation (2.6) has some advantages over the dual approach. First, $V_0(X_0)$ depends on v_i monotonically, i.e., if we have two lower bounds v and $\widetilde{v}$ such that $v_i(X_i) \le \widetilde{v}_i(X_i)$ for all i then $V_0(X_0) \ge \widetilde{V}_0(X_0)$. This immediately follows from the first line in (2.14). This is not always the case for the dual method. Indeed, with three exercises ($\mathcal{I} = 2$) formula (2.11) gives

$$V_0^D = E\max\{f_0, E(v_1|\mathcal{F}_0) + u_1 - v_1\}.$$

Consider the case when the probability of event $A := \{Ev_1 - u_1 - v_1 \ge f_0\}$ is positive and $v_1 < u_1 - \theta$ with some constant $\theta > 0$.. Then taking $\widetilde{v}_1 = v_1 + \theta/2$ on A and $\widetilde{v}_1 = v_1 + \theta$ outside A, we obtain

$$\widetilde{V}_0^D := E\max\{f_0, E(\widetilde{v}_1|\mathcal{F}_0) + u_1 - \widetilde{v}_1\} > V_0^D,$$

though $\widetilde{v}_1 > v_1$. Second, adaptive local lower bounds of the form

$$v_i(x) = \max_{1\le k\le l} v_i^k(x), \quad i = 0, \dots, \mathcal{I}-1,$$

where $v_1(x), \dots, v_l(x)$ are lower bounds at x ordered according to their complexity and l may depend on x, can be used to construct $V_0(X_0)$ (see [4]). It is also worthwhile mentioning that our approach allows us to construct lower bounds using upper ones.

2.5 Bermudan options

As before, we consider the discrete-time model

$$(B_i, X_i) = (B_i, X_i^1, ..., X_i^d), \quad i = 0, 1, ..., \mathcal{I}.$$

However, now the holder can exercise his right only at time belonging to the set of stopping times $S = \{s_1, ..., s_l\}$ within $\{0, 1, ..., \mathcal{I}\}$, where $s_l = \mathcal{I}$. The price $u_i(X_i)$ of the Bermudan option is given by

$$u_i(X_i) = \sup_{\tau\in\mathcal{T}_{S\cap[i,\mathcal{I}]}} B_i E\left(\frac{f_\tau(X_\tau)}{B_\tau}|\mathcal{F}_i\right),$$

where $\mathcal{T}_{S\cap[i,\mathcal{I}]}$ is the set of stopping times τ taking values in $\{s_1, ..., s_l\} \cap \{i, i+1, ..., \mathcal{I}\}$.

The value process u_i is determined as follows:

$$u_{\mathcal{I}}(x) = f_{\mathcal{I}}(x),$$
$$u_i(x) = \begin{cases} \max\left\{f_i(x),\ B_i E\left(\frac{u_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right)\right\},\ i \in S, \\ B_i E\left(\frac{u_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right),\ i \notin S. \end{cases}$$

Thus, we obtain that the Bermudan option is equivalent to the European option with payoff function $f_i(x)$ and consumption process γ_i defined by

$$\gamma_i(x) = \begin{cases} \left[f_i(x) - B_i E\left(\frac{u_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right)\right]^+,\ i \in S \\ 0,\ i \notin S. \end{cases}$$

From here, all the results for discrete-time American options obtained in this section can be carried over to the Bermudan options. For example, if $v_i(x)$ is a lower bound of the true option price $u_i(x)$, the price $V_i(x)$ of the European option with payoff function $f_{\mathcal{I}}(x)$ and consumption process

$$\gamma_{i,v}(x) = \begin{cases} \left[f_i(x) - B_i E\left(\frac{v_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right)\right]^+,\ i \in S, \\ 0,\ i \notin S, \end{cases}$$

is an upper bound: $V_i(x) \geq u_i(x)$.

3 Optimal stopping times and algorithms with lower continuation values

The samples with optimal stopping times have been first introduced in [23] (see also [12]). In this section we first recall some basic relations for optimal stopping times in the form suitable for our purposes. Then we show that subsequent estimating these times amounts to evaluation of continuation values by regression. There are many nonparametric regression methods available (see, e.g., [16]). In Subsection 3.3 we propose some algorithms based both on local modelling and least squares estimation. Using the regression approach for pricing American options, we construct not only an estimate for the continuation value but also an upper consumption process.

3.1 Basic relations for optimal stopping times

The optimal stopping time $\tau^{i,x} = \tau^{t_i,x}$ depends on the initial position (t_i, x). It is defined recurrently by the dynamic programming principle in the following way. We set

$$\begin{aligned} \tau^{\mathcal{I},x} &= \tau^{T,x} = T, \\ \tau^{i,x} &= t_i \chi_{\{C_i(x) \le f_i(x)\}} + \tau^{i+1, X_{i+1}^{i,x}} \chi_{\{C_i(x) > f_i(x)\}} \\ &= t_i \chi_{\{u_i(x) = f_i(x)\}} + \tau^{i+1, X_{i+1}^{i,x}} \chi_{\{u_i(x) > f_i(x)\}}, \\ i &= \mathcal{I}-1, ..., 0. \end{aligned} \tag{3.1}$$

Thus, for any position (t_i, x), the optimal stopping time $\tau^{i,x}$ is either equal to t_i : $\tau^{i,x} = t_i$, or $\tau^{i,x} > t_i$. Hence

$$\tau^{i,x} = \tau^{i+1, X_{i+1}^{i,x}}, \text{ if } \tau^{i,x} > t_i. \tag{3.2}$$

It is also clear that (t_i, x) is a stopping point (i.e., $\tau^{i,x} = t_i$) if and only if $(t_i, x) \in \mathcal{E}$ (i.e., (t_i, x) belongs to the exercise region). The instant $\tau^{i,x}$ is the first one at which the trajectory $(t_j, X_j^{i,x})$ either enters $\mathcal{E}$ during $i \le j \le \mathcal{I}-1$ or stops at the final time $\mathcal{I}$. So, $(\tau^{i,x}, X_{\tau^{i,x}}^{i,x}) \in \mathcal{E}$ (see (2.2)). Let us give some recurrence relations for $u_i(x)$ and $C_i(x)$:

$$u_i(X_i) = \max\{f_i(X_i), C_i(X_i)\}, \; u_{\mathcal{I}}(x) = f(x), \tag{3.3a}$$

$$C_i(X_i) = \frac{B_i}{B_{i+1}} E(u_{i+1}(X_{i+1})|X_i), \; C_{\mathcal{I}}(x) = f(x), \tag{3.3b}$$

$$C_i(X_i) = \frac{B_i}{B_{i+1}} E(\max\{f_{i+1}(X_{i+1}), C_{i+1}(X_{i+1})\}|X_i), \tag{3.3c}$$

$$u_i(X_i) = \max\{f_i(X_i), \frac{B_i}{B_{i+1}} E(u_{i+1}(X_{i+1})|X_i)\}. \tag{3.3d}$$

We note that

$$u_{i+1}(X_{i+1}) = B_{i+1} E\left(\frac{f_\tau(X_\tau^{t_{i+1}, X_{i+1}})}{B_\tau} | X_{i+1} \right), \tag{3.4}$$

$$\begin{aligned} E(u_{i+1}(X_{i+1})|X_i) &= E\left(B_{i+1} E\left(\frac{f_\tau(X_\tau^{t_{i+1}, X_{i+1}})}{B_\tau} | \mathcal{F}_{i+1} \right) | \mathcal{F}_i \right) \\ &= B_{i+1} E\left(\frac{f_\tau(X_\tau^{t_{i+1}, X_{i+1}})}{B_\tau} | X_i \right), \end{aligned} \tag{3.5}$$

where

$$\tau = \tau^{t_{i+1}, X_{i+1}}.$$

Hence, due to (3.3b), we get

$$C_i(X_i) = B_i E\left(\frac{f_\tau(X_\tau^{t_{i+1}, X_{i+1}})}{B_\tau} | X_i \right). \tag{3.6}$$

We emphasize that for any stopping time $\widetilde{\tau} \geq t_{i+1}$ the function

$$v_{i+1}(x) = B_{i+1} E\left(\frac{f_{\widetilde{\tau}}(X_{\widetilde{\tau}}^{t_{i+1},x})}{B_{\widetilde{\tau}}}\right) \tag{3.7}$$

is a lower bound for $u_{i+1}(x)$.

Since

$$C_i(x) = \sup_{\tau \in \mathcal{T}_{i+1,\mathcal{I}}} B_i E\left(\frac{f_\tau(X_\tau^{t_{i+1},X_{i+1}})}{B_\tau} | X_i = x\right) = \sup_{\tau \in \mathcal{T}_{i+1,\mathcal{I}}} B_i E\left(\frac{f_\tau(X_\tau^{t_i,x})}{B_\tau}\right), \tag{3.8}$$

the function

$$c_i(x) = B_i E\left(\frac{f_{\widetilde{\tau}}(X_{\widetilde{\tau}}^{t_i,x})}{B_{\widetilde{\tau}}}\right) \tag{3.9}$$

is a lower continuation value for any stopping time $\widetilde{\tau} \geq t_{i+1}$.

3.2 Estimating optimal stopping times

Considering $C_i(x)$ as a regression function (see (3.6)), it is natural to introduce (after [23] and [12]) the sample

$$\begin{aligned}({}_mX_i,\ \frac{B_i}{B_\tau} f_\tau({}_mX_\tau^{t_{i+1},\ {}_mX_{i+1}})) \ &= \ ({}_mX_i,\ \frac{B_i}{B_\tau} f_\tau({}_mX_\tau^{t_i,\ {}_mX_i})), \\ &\tau = \tau^{t_{i+1},\ {}_mX_{i+1}},\ m = 1, ..., M,\end{aligned} \tag{3.10}$$

from $(X_i,\ \frac{B_i}{B_\tau} f_\tau(X_\tau^{t_{i+1},X_{i+1}})) = (X_i,\ \frac{B_i}{B_\tau} f_\tau(X_\tau^{t_i,X_i}))$, where $\tau = \tau^{t_{i+1},X_{i+1}}$.

We are about to use (3.10) for subsequent constructing an estimate $\widehat{\tau}^{t_i,\ {}_mX_i}$ for the optimal stopping time $\tau^{t_i,\ {}_mX_i}$. Clearly, $\tau^{\mathcal{I},\ {}_mX_\mathcal{I}} = \widehat{\tau}^{\mathcal{I},\ {}_mX_\mathcal{I}} = \mathcal{I}$. Let $\tau^{t_{i+1},\ {}_mX_{i+1}}$, $i = \mathcal{I}-1, ..., 1$, (in reality $\widehat{\tau}^{t_{i+1},\ {}_mX_{i+1}}$) be already estimated. Using the sample (3.10) at the step t_i, we evaluate $C_i({}_mX_i)$ by regression. Let $\widehat{C}_i({}_mX_i)$ be an estimate of $C_i({}_mX_i)$. If $f_i({}_mX_i) \geq \widehat{C}_i({}_mX_i)$ then $\widehat{\tau}^{t_i,\ {}_mX_i} = t_i$, otherwise $\widehat{\tau}^{t_i,\ {}_mX_i} = \widehat{\tau}^{t_{i+1},\ {}_mX_{i+1}}$ (see (3.1)). As a result, we obtain a sample like (3.10) at the previous time step t_{i-1}:

$$\begin{aligned}({}_mX_{i-1},\ \frac{B_{i-1}}{B_\tau} f_\tau({}_mX_\tau^{t_i,\ {}_mX_i})) \ &= \ ({}_mX_{i-1},\ \frac{B_{i-1}}{B_\tau} f_\tau({}_mX_\tau^{t_{i-1},\ {}_mX_{i-1}})), \\ &\tau = \tau^{t_i,\ {}_mX_i},\ m = 1, ..., M.\end{aligned} \tag{3.11}$$

This allows us to construct the estimate $\widehat{C}_{i-1}({}_mX_{i-1})$ of $C_{i-1}({}_mX_{i-1})$ and the estimate $\widehat{\tau}^{t_{i-1},\ {}_mX_{i-1}}$ of $\tau^{t_{i-1},\ {}_mX_{i-1}}$ and so on. Upon estimating $\tau^{t_1,\ {}_mX_1}$, we can evaluate $u_0(X_0)$ via

$$u_0(X_0) = \max\{f_0(X_0), \frac{1}{B_1} E(u_1(X_1^{t_0,X_0})\} = \max\left\{f_0(X_0), E\left(\frac{f_\tau(X_\tau^{t_1,X_1})}{B_\tau}\right)\right\},\ \tau = \tau^{t_1,X_1}. \tag{3.12}$$

So, we construct continuation values and stopping times simultaneously by the backward procedure and our main problem is to evaluate the continuation value $C_i({}_mX_i)$

using the sample (3.10). To this aim, we use nonparametric regression methods. In the next subsection we propose some algorithms based both on local modelling and least squares estimation.

The most appropriate are methods for which the estimate $\widehat{C}_i({}_mX_i)$ is a lower continuation value. If the payoff at $(t_i,\ {}_mX_i)$ is less than or equal to a lower continuation value, then first, the position $(t_i,\ {}_mX_i)$ belongs to the continuation region (consequently, it is natural to take $\widehat{\tau}^{t_i,\ {}_mX_i} = \widehat{\tau}^{t_{i+1},\ {}_mX_{i+1}}$) and, second, the consumption process at $(t_i,\ {}_mX_i)$ is equal to zero. Otherwise the position $(t_i,\ {}_mX_i)$ can belong either to the exercise region or to the continuation region. In the latter case we compute the upper consumption process at $(t_i,\ {}_mX_i)$ as a difference between the payoff and the lower continuation value and set $\widehat{\tau}^{t_i,\ {}_mX_i} = t_i$. As a result all positions $(t_i,\ {}_mX_i)$ are equipped with the stopping times and the upper consumption processes. In such a situation we are able to construct both lower and upper bounds for the price of the option under consideration.

3.3 Algorithms with the local Monte Carlo approach

For every position $(t_i,\ {}_mX_i)$, $m = 1, ...M$, let us construct $N = N_{i,m}$ additional independent trajectories on $[t_i, t_{i+1}]$, i.e., the trajectories with the length of one step. At the instant t_{i+1} we obtain $N+1$ points ${}_nX_{t_{i+1}}^{t_i,\ {}_mX_i}$, $n = 0, 1, ..., N$, where we put ${}_0X_{t_{i+1}}^{t_i,\ {}_mX_i}$ $=_m X_{i+1}$. Introduce the notation ${}_{m,n}X_{i+1} :=_n X_{t_{i+1}}^{t_i,\ {}_mX_i}$, $\tau_{m,n} :=_n \tau^{t_{i+1},\ {}_{m,n}X_{i+1}}$. It follows from the semigroup property for the Markov chain $X_{t_{i+j}}^{t_i,\ x}$ that if $\tau^{t_i,\ {}_mX_i} \geq t_{i+1}$, then $\tau^{t_i,\ {}_mX_i} = \tau^{t_{i+1},\ X_{i+1}(t_i,\ {}_mX_i)}$, where the notation $X_{i+1}(t_i,\ {}_mX_i) = X_{t_{i+1}}^{t_i,\ {}_mX_i}$ is used. This is true for the n-th independent copy ${}_nX_{t_{i+1}}^{t_i,\ {}_mX_i}$ of $X_{t_{i+1}}^{t_i,\ {}_mX_i}$ as well, i.e., ${}_n\tau^{t_i,\ {}_mX_i} =_n \tau^{t_{i+1},\ {}_{m,n}X_{i+1}} = \tau_{m,n}$, if $\tau^{t_i,\ {}_mX_i} \geq t_{i+1}$. Due to (3.6), we have

$$C_i({}_mX_i) = B_i E\left(\frac{f_\tau(X_\tau^{t_{i+1},X_{i+1}})}{B_\tau}|X_i =_m X_i\right) \simeq \frac{B_i}{N+1}\sum_{n=0}^{N}\frac{f_{\tau_{m,n}}(X_{\tau_{m,n}}^{t_{i+1},\ {}_{m,n}X_{i+1}})}{B_{\tau_{m,n}}}. \tag{3.13}$$

Let us stress that the sum in (3.13) is an estimate of $C_i({}_mX_i)$ in a theoretical sense only because we do not simulate the trajectory $X_{t_j}^{t_{i+1},\ {}_{m,n}X_{i+1}}$ for $t_j > t_{i+1}$. That is why for every point ${}_{m,n}X_{i+1} =_n X_{t_{i+1}}^{t_i,\ {}_mX_i}$, we find the nearest one among ${}_kX_{i+1}$, $k = 1, ...M$, denoted by ${}_{k(m,n)}X_{i+1}$. For the position $(t_{i+1},\ {}_{k(m,n)}X_{i+1})$, the estimate $\widehat{\tau}_{k(m,n)}$ of the optimal stopping time $\tau^{t_{i+1},\ {}_{k(m,n)}X_{i+1}}$ is known. To avoid confusion, let us emphasize that the points ${}_{m,n}X_{i+1}$ lie on the trajectories starting from the same position $(t_i,\ {}_mX_i)$ and for the positions $(t_{i+1},\ {}_{m,n}X_{i+1})$ estimates of the optimal stopping times are in general unknown, while the points ${}_{k(m,n)}X_{i+1}$ lie on the trajectories which have different starting positions $(t_i,\ {}_{k(m,n)}X_i)$.

For the continuation value $C_i({}_mX_i)$, we introduce the estimate

$$\widehat{C}_i({}_mX_i) = \frac{B_i}{N+1}\sum_{n=0}^{N}\frac{f_{\widehat{\tau}_{k(m,n)}}(X_{\widehat{\tau}_{k(m,n)}}^{t_{i+1},\ {}_{k(m,n)}X_{i+1}})}{B_{\widehat{\tau}_{k(m,n)}}}. \tag{3.14}$$

In distinction to (3.13), this estimate is simulated. We intend to prove that in a sense $\widehat{C}_i({}_mX_i)$ is a low continuation value. To this aim, we consider an auxiliary low continuation value $\widetilde{C}_i({}_mX_i)$ (which is not simulated) and then prove that $\widehat{C}_i({}_mX_i)$ is close to $\widetilde{C}_i({}_mX_i)$.

For any point $X_{i+1} = X_{t_{i+1}}^{t_i,\ {}_mX_i}$, one can define the stopping time $\widetilde{\tau} = \widetilde{\tau}(X_{i+1}) \geq t_{i+1}$ analogously to $\widehat{\tau}_{k(m,n)}$, i.e., first, we find the nearest point to X_{i+1} among ${}_kX_{i+1}$, $k = 1, ...M$, say ${}_{\widetilde{k}}X_{i+1}$, and, second, for the position $(t_{i+1},\ {}_{\widetilde{k}}X_{i+1})$ we know the estimate $\widehat{\tau}_{\widetilde{k}}$ of the optimal stopping time $\tau^{t_{i+1},\ {}_{\widetilde{k}}X_{i+1}}$, which we take as $\widetilde{\tau} : \widetilde{\tau} = \widetilde{\tau}(X_{i+1}) = \widehat{\tau}_{\widetilde{k}}$. Clearly, for the points ${}_{m,n}X_{i+1}$ this stopping time $\widetilde{\tau} = \widetilde{\tau}({}_{m,n}X_{i+1}) := \widetilde{\tau}_{m,n}$ coincides with $\widehat{\tau}_{k(m,n)}$. Introduce

$$\widetilde{C}_i(x) = B_i E\left(\frac{f_{\widetilde{\tau}}(X_{\widetilde{\tau}}^{t_{i+1},X_{i+1}})}{B_{\widetilde{\tau}}}|X_i = x\right).$$

It follows from (3.8) and (3.9) that

$$C_i(x) = \widetilde{C}_i(x) + r_i(x), \tag{3.15}$$

where $r_i(x) \geq 0$, i.e., $\widetilde{C}_i(x)$ is a lower continuation value at the position $(t_i,\ x)$. Further, we have

$$\begin{aligned}\widetilde{C}_i({}_mX_i) &= \frac{B_i}{N+1}\sum_{n=0}^{N}\frac{f_{\widetilde{\tau}_{m,n}}(X_{\widetilde{\tau}_{m,n}}^{t_{i+1},\ {}_{m,n}X_{i+1}})}{B_{\widetilde{\tau}_{m,n}}} + \alpha_i({}_mX_i) \\ &= \frac{B_i}{N+1}\sum_{n=0}^{N}\frac{f_{\widehat{\tau}_{k(m,n)}}(X_{\widehat{\tau}_{k(m,n)}}^{t_{i+1},\ {}_{m,n}X_{i+1}})}{B_{\widehat{\tau}_{k(m,n)}}} + \alpha_i({}_mX_i),\end{aligned} \tag{3.16}$$

where $\alpha_i({}_mX_i)$ is the Monte Carlo error which becomes small with increasing N. Let us note that in general the points $X_{\widetilde{\tau}_{m,n}}^{t_{i+1},\ {}_{m,n}X_{i+1}}$ do not belong to the considered sample of M independent trajectories all starting from the initial point (t_0, X_0). That is why the sum in (3.16) cannot be taken as an estimate for the continuation value $C_i({}_mX_i)$.

Let us note that in (3.14) and in (3.16) we consider the trajectories $X_s^{t_{i+1},\ {}_{k(m,n)}X_{i+1}}$ and $X_s^{t_{i+1},\ {}_{m,n}X_{i+1}}$ starting from the different positions $(t_{i+1},\ {}_{k(m,n)}X_{i+1})$ and $(t_{i+1},\ {}_{m,n}X_{i+1})$ but with the same sources of randomness. If M is large, the points ${}_{m,n}X_{i+1}$ and ${}_{k(m,n)}X_{i+1}$ are close to each other and we get

$$\begin{aligned}\widehat{C}_i({}_mX_i) &= \frac{B_i}{N+1}\sum_{n=0}^{N}\frac{f_{\widehat{\tau}_{k(m,n)}}(X_{\widehat{\tau}_{k(m,n)}}^{t_{i+1},\ {}_{m,n}X_{i+1}})}{B_{\widehat{\tau}_{k(m,n)}}} - \beta_i({}_mX_i) \\ &= \widetilde{C}_i({}_mX_i) - \alpha_i({}_mX_i) - \beta_i({}_mX_i),\end{aligned} \tag{3.17}$$

where the approximation error β_i is small.

From (3.15) we obtain

$$\widehat{C}_i({}_mX_i) = C_i({}_mX_i) + \rho_i({}_mX_i) - r_i({}_mX_i), \tag{3.18}$$

where $\rho_i = -\alpha_i - \beta_i$.

Because $r_i \geq 0$, the quantity $C_i({}_mX_i) - r_i({}_mX_i)$ is a lower continuation value. The estimate $\widehat{C}_i({}_mX_i)$ differs from this quantity for ρ_i, i.e., $\widehat{C}_i({}_mX_i)$ is a lower continuation value within the accuracy ρ_i. We emphasize that ρ_i becomes small with increasing N and M. Thus the following proposition is justified.

Proposition 3.1. *The estimate $\widehat{C}_i({}_mX_i)$ is a lower continuation value within the accuracy depending on N (the accuracy determined by the Monte Carlo error) and M (the accuracy determined by the approximation error).*

Corollary 3.2. *Consider the consumption*

$$\widehat{\gamma}_i({}_mX_i) = [f_i({}_mX_i) - \widehat{C}_i({}_mX_i)]^+. \tag{3.19}$$

Since

$$\widehat{\gamma}_i({}_mX_i) = [f_i({}_mX_i) - C_i({}_mX_i) + r_i({}_mX_i) - \rho_i({}_mX_i)]^+$$

and $\gamma_i({}_mX_i) = [f_i({}_mX_i) - C_i({}_mX_i)]^+$, we have

$$\gamma_i({}_mX_i) \leq \widehat{\gamma}_i({}_mX_i), \ \text{if } r_i \geq \rho_i, \tag{3.20}$$
$$[\gamma_i({}_mX_i) - \rho_i({}_mX_i) + r_i({}_mX_i)]^+ = \widehat{\gamma}_i({}_mX_i) \leq \gamma_i({}_mX_i), \ \text{if } \rho_i > r_i.$$

We see that $\widehat{\gamma}_i({}_mX_i)$ is an upper consumption in the most typical case $r_i \geq \rho_i$, otherwise $\widehat{\gamma}_i({}_mX_i)$ is close to $\gamma_i({}_mX_i)$. Thus, $\widehat{\gamma}_i({}_mX_i)$ is an upper consumption within the accuracy depending on M and N.

3.4 Independence of estimates of future information

The continuation value $C_i({}_mX_i)$ due to (3.13) does not depend on any future information. The right-hand side of (3.13) is close to $C_i({}_mX_i)$ for large N and this implies closeness of the estimate $\widehat{C}_i({}_mX_i)$ to $C_i({}_mX_i)$ for large M. This fact is the most important one for the quality of a consumption's estimate. As to the independence of future information, it is intuitively clear that, e.g., for large N the right-hand side of (3.13) can only weakly depend on the future behavior of the trajectories $({}_{t_j,m}X_j)$ for $t_j > t_i$. At the same time, it is not difficult to rigorously construct independent estimates for the continuation values (of course, with higher simulation cost). To this aim, along with the old set of trajectories, let us simulate $\breve{M}$ new independent trajectories $({}_{t_i,m}\breve{X}_i)$, $i = 0, ..., \mathcal{I}-1$, $m = 1, ..., \breve{M}$, all starting from the point (t_0, X_0). Further, the points ${}_{m,n}\breve{X}_{i+1}$, $m = 1, ..., \breve{M}$, $n = 1, ..., \breve{N}$, are simulated instead of ${}_{m,n}X_{i+1}$, and then for every point ${}_{m,n}\breve{X}_{i+1}$ we find the nearest point ${}_{\breve{k}(m,n)}X_{i+1}$among ${}_kX_{i+1}$, $k = 1, ..., M$ (hence, we find $\widehat{\tau}_{\breve{k}(m,n)}$). Clearly, the estimate

$$\breve{C}_i({}_m\breve{X}_i) = \frac{B_i}{\breve{N}} \sum_{n=1}^{\breve{N}} \frac{f_{\widehat{\tau}_{\breve{k}(m,n)}}(X^{t_{i+1},\ {}_{\breve{k}(m,n)}X_{i+1}}_{\widehat{\tau}_{\breve{k}(m,n)}})}{B_{\widehat{\tau}_{\breve{k}(m,n)}}} \tag{3.21}$$

does not depend on the future behavior of the trajectories $({}_{t_j,m}\breve{X}_j)$ for $t_j > t_i$. So, every new position $({}_{t_i,m}\breve{X}_i)$ is equipped with the lower continuation value $\breve{C}_i({}_m\breve{X}_i)$. It

can be equipped with the estimate $\breve{\tau}^{t_i,{}_m\breve{X}_i}$ of the optimal stopping time $\tau^{t_i,{}_m\breve{X}_i}$ as well using the rule: $\breve{\tau}^{\mathcal{I},{}_m\breve{X}_{\mathcal{I}}} = \mathcal{I}$ for $i = \mathcal{I}$; $\breve{\tau}^{t_i,{}_m\breve{X}_i} = t_i$ if $f_i({}_m\breve{X}_i) \geq \breve{C}_i({}_m\breve{X}_i)$, otherwise $\breve{\tau}^{t_i,{}_m\breve{X}_i} = \breve{\tau}^{t_{i+1},{}_m\breve{X}_{i+1}}$ for $i = \mathcal{I}-1,...,1$. This allows us to effectively find lower and upper bounds for the price of the option at the initial position (t_0, X_0) without using any future information (see Section 4). Nevertheless, we prefer the estimate (3.14), which is of lower simulation cost and which uses very little future information. Additionally, let us emphasize that the estimates like (3.14) give in practice almost the same results as their counterparts of the form (3.21) (see Section 5).

3.5 Algorithms with the local Monte Carlo approach, continuation

In the estimate (3.14) we use the points ${}_{k(m,n)}X_{i+1}$ which are chosen among ${}_mX_{i+1}$, $m = 1,\dots,M$, as the nearest ones to ${}_{m,n}X_{i+1}$. Now for every point ${}_{m,n}X_{i+1} =_n X_{t_{i+1}}^{t_i,\ {}_mX_i}$ let us find a few (say $K_{m,n}$) nearest ones among ${}_mX_{i+1}$. Let us denote them by ${}_{k[m,n]}X_{i+1}$, $k = 1,...,K_{m,n}$ (in contrast to $k(m,n)$, the function $k[m,n]$ is a multifunction). The estimates $\widehat{\tau}_{k[m,n]}$ of the optimal stopping times $\tau_{k[m,n]} := \tau^{t_{i+1},\ {}_{k[m,n]}X_{i+1}}$ are known. Then the following expression

$$v_{i+1}({}_nX_{t_{i+1}}^{t_i,\ {}_mX_i}) = \frac{B_{i+1}}{K_{m,n}} \sum_{k=1}^{K_{m,n}} \frac{f(X_{\widehat{\tau}_{k[m,n]}}^{t_{i+1},\ {}_{k[m,n]}X_{i+1}})}{B_{\widehat{\tau}_{k[m,n]}}} \tag{3.22}$$

is a lower bound for $u_{i+1}(x)$ at the position $(t_{i+1},\ {}_nX_{t_{i+1}}^{t_i,\ {}_mX_i})$ (of course, within accuracy of the approximation).

Clearly,

$$\widehat{C}_i({}_mX_i) = \frac{B_i}{B_{i+1}} \cdot \frac{1}{N+1} \sum_{n=0}^{N} v_{i+1}({}_nX_{t_{i+1}}^{t_i,\ {}_mX_i}) = \frac{B_i}{N+1} \sum_{n=0}^{N} \frac{1}{K_{m,n}} \sum_{k=1}^{K_{m,n}} \frac{f(X_{\widehat{\tau}_{k[m,n]}}^{t_{i+1},\ {}_{k[m,n]}X_{i+1}})}{B_{\widehat{\tau}_{k[m,n]}}} \tag{3.23}$$

is a lower continuation value at $(t_i,\ {}_mX_i)$ (of course, within accuracy depending on M and N). The estimate (3.14) is the particular case of (3.23), when $K_{m,n} = 1$. Let us note that for the estimate (3.23), analogues of Proposition 3.1, Corollary 3.2, and Subsection 3.4 hold as well.

3.6 Algorithms with k-NN estimates

In the previous algorithms we construct $N_{i,m}$ additional trajectories for every point ${}_mX_i$, $m = 1,...,M$. Let us consider $N = N_{i,m}$ nearest points ${}_{m,1}X_i,...,{}_{m,N}X_i$ to the point ${}_mX_i$ instead of constructing the additional trajectories. All the points ${}_{m,1}X_i,...,{}_{m,N}X_i$ belong to the set $\{\ {}_mX_i,\ m = 1,...,M\}$. We have ${}_{m,n}X_{i+1}^{(t_i,\ {}_{m,n}X_i)} =_{m,n} X_{i+1}$, $n = 0,1,...,N$, ${}_{m,0}X_i =_m X_i$, ${}_{m,0}X_{i+1} =_m X_{i+1}$, with known $\widehat{\tau}_{m,n} = \widehat{\tau}^{t_{i+1},\ {}_{m,n}X_{i+1}}$ and $f(X_{\widehat{\tau}^{t_{i+1},\ {}_{m,n}X_{i+1}}}^{(t_{i+1},\ {}_{m,n}X_{i+1})})$ (let us note that we use another notation in this subsection and, in

particular, we emphasize that the points ${}_{m,n}X_{i+1}$ belong to the set $\{ {}_mX_{i+1},\ m = 1,...M\}$). Then analogously to (3.14), we evaluate

$$\widehat{C}_i({}_mX_i) = \frac{B_i}{N+1}\sum_{n=0}^{N}\frac{f_{\widehat{\tau}_{m,n}}(X_{\widehat{\tau}_{m,n}}^{t_{i+1},\ {}_{m,n}X_{i+1}})}{B_{\widehat{\tau}_{m,n}}}. \tag{3.24}$$

To get an analogue of (3.23), let us find a few (say $K_{m,n}$) points among ${}_mX_{i+1}$, $m = 1,...M$, which are nearest to ${}_{m,n}X_{i+1} = {}_{m,n}\ X_{i+1}^{(t_i,\ {}_{m,n}X_i)}$. Denote them by ${}_{m,n,k}X_{i+1}$, $k = 1,...,K_{m,n}$. Then

$$\widehat{C}_i({}_mX_i) = \frac{B_i}{N+1}\sum_{n=0}^{N}\frac{1}{K_{m,n}}\sum_{k=1}^{K_{m,n}}\frac{f(X_{\widehat{\tau}_{m,n,k}}^{t_{i+1},\ {}_{m,n,k}X_{i+1}})}{B_{\widehat{\tau}_{m,n.k}}}, \tag{3.25}$$

where $\widehat{\tau}_{m,n,k}$ are known estimates of the optimal stopping times $\tau_{m,n,k} := \tau^{t_{i+1},\ {}_{m,n,k}X_{i+1}}$. Note that ${}_{m,n,k}X_{i+1}$ in (3.25) are different from ${}_{k[m,n]}X_{i+1}$ in (3.23).

In the case of (3.25), analogues of Proposition 3.1, Corollary 3.2, and Subsection 3.4 hold as well.

Remark 3.1. The k-NN estimates belong to the class of local averaging estimates (see [16]). The proper choice of $K_{m,n}$ is important and is, e.g., discussed in [7]. One can use other estimates of this class, e.g., kernel estimates and local polynomial kernel estimates. Note that the latter type of estimates can be helpful for estimating deltas.

3.7 Linear regression

Regression-based methods approximate the continuation value using a basis function expansion:

$$C_i(x) \approx \sum_{r=1}^{K}\beta_{ir}\psi_r(x), \quad i = 0,1,\dots,\mathcal{I}-1,$$

where $\{\psi_r(x)\}_{r=1}^K$ is a set of basis functions each mapping $\mathbf{X}$ to $\mathbf{R}$. Following the notation in [15], we have

$$C_i(x) \approx \beta_i^\top\psi(x)$$

with

$$\beta_i^\top = (\beta_{i1},\dots,\beta_{iK}), \quad \psi(x) = (\psi_1(x),\dots,\psi_K(x))^\top.$$

The vector β_i can be estimated using the sample

$$({}_mX_i,\ \frac{B_i}{B_{\widehat{\tau}_m}}f_{\widehat{\tau}_m}({}_mX_{\widehat{\tau}_m}^{t_{i+1},\ {}_mX_{i+1}})), \quad \widehat{\tau}_m = \widehat{\tau}^{t_{i+1},\ {}_mX_{i+1}}, \quad m = 1,\dots,M,$$

as

$$\widehat{\beta}_i = \widehat{A}_\psi^{-1}\widehat{\alpha}_{\psi V}.$$

Here $\widehat{A}_\psi$ is the $K\times K$ matrix with qr entry

$$\frac{1}{M}\sum_{m=1}^{M}\psi_q({}_mX_i)\psi_r({}_mX_i)$$

and $\widehat{\alpha}_{\psi V}$ is the K-vector with r-th entry

$$\frac{1}{M}\sum_{m=1}^{M}\psi_r({}_mX_i)\frac{B_i f_{\widehat{\tau}_m}(X_{\widehat{\tau}_m}^{t_{i+1},\,{}_mX_{i+1}})}{B_{\widehat{\tau}_m}}.$$

The estimate $\widehat{\beta}_i$ then defines an estimate

$$\widehat{C}_i(x)=\widehat{\beta}_i^{\top}\psi(x)$$

of the continuation value at an arbitrary point x in the state space $\mathbf{X}$. Now, if $f_i({}_mX_i)\geq \widehat{C}_i({}_mX_i)$ then $\widehat{\tau}^{t_i,\,{}_mX_i}=t_i$, otherwise $\widehat{\tau}^{t_i,\,{}_mX_i}=\widehat{\tau}^{t_{i+1},\,{}_mX_{i+1}}$ (see (3.1)). As a result, at the step t_{i-1} we obtain the sample

$$\begin{aligned}({}_mX_{i-1},\ \frac{B_{i-1}}{B_{\widehat{\tau}_m}}f_{\widehat{\tau}_m}({}_mX_{\widehat{\tau}_m}^{t_i,\,{}_mX_i}))\ &=\ ({}_mX_{i-1},\ \frac{B_{i-1}}{B_{\widehat{\tau}_m}}f_{\widehat{\tau}_m}({}_mX_{\widehat{\tau}_m}^{t_{i-1},\,{}_mX_{i-1}})),\\ &\quad\ \widehat{\tau}_m=\widehat{\tau}^{t_i,\,{}_mX_i},\ m=1,...,M.\end{aligned}$$

Proposition 3.3. *The estimate*

$$\widehat{C}_i({}_mX_i)=\widehat{\beta}_i^{\top}\psi({}_mX_i) \tag{3.26}$$

is a lower continuation value within the accuracy depending on K and M.

Proof. Having $\widehat{C}_j(x),\ x\in\mathbf{X},\ j=0,...,\mathcal{I}-1$, one can define a stopping time $\widetilde{\tau}$ for every trajectory $X_{t_j}^{t_i,\,x},\ j=i,...,\mathcal{I}$, in the following way. If $\widehat{C}_i(x)\leq f_i(x)$, then we put $\widehat{\tau}^{t_i,\,x}=t_i$. If $\widehat{C}_i(x)>f_i(x)$, then we put $\widehat{\tau}^{t_i,\,x}>t_i$. Further, if $\widehat{C}_{i+1}(X_{t_{i+1}}^{t_i,\,x})\leq f_{i+1}(X_{t_{i+1}}^{t_i,\,x})$, then we put $\widehat{\tau}^{t_i,\,x}=t_{i+1}$, and so on. If $\widehat{C}_j(X_{t_j}^{t_i,\,x})>f_j(X_{t_j}^{t_i,\,x})$ for all $j=i,...,\mathcal{I}-1$, then we put $\widehat{\tau}^{t_i,\,x}=\mathcal{I}$. Clearly, $\widetilde{\tau}^{t_i,\,{}_mX_i}=\widehat{\tau}^{t_i,\,{}_mX_i},\ m=1,...,M$, i.e., $\widetilde{\tau}$ is an extension of $\widehat{\tau}$. Let us introduce the value

$$\widetilde{C}_i(x)=B_iE\left(\frac{f_{\widetilde{\tau}}(X_{\widetilde{\tau}}^{t_{i+1},X_{i+1}})}{B_{\widetilde{\tau}}}|X_i=x\right),\ \widetilde{\tau}=\widetilde{\tau}^{t_{i+1},\,X_{i+1}}. \tag{3.27}$$

Due to (3.8) and (3.9), $\widetilde{C}_i(x)$ is a lower continuation value, i.e.,

$$\widetilde{C}_i(x)=C_i(x)-r_i(x), \tag{3.28}$$

where $r_i(x)\geq 0$. But in the conditional expectation (3.27), $\widehat{C}_i(x)$ can be considered as an estimate by the linear regression method. Therefore

$$\widetilde{C}_i(x)=\widehat{C}_i(x)+\alpha_i(x), \tag{3.29}$$

where $\alpha_i(x)$ is the regression error which depends on K and M. It follows from (3.28) and (3.29):

$$\widehat{C}_i({}_mX_i)=C_i({}_mX_i)-\alpha_i({}_mX_i)-r_i({}_mX_i). \tag{3.30}$$

Proposition 3.3 is proved. □

Let us note that for the estimate (3.23), analogues of Corollary 3.2, and Subsection 3.4 hold as well.

Remark 3.2. In fact, Proposition 3.3 states that $\widehat{C}_i({}_mX_i) - \mid \alpha_i({}_mX_i) \mid$ is a lower continuation value (see also (3.30)). I.e., the proposition is true even if the error $\alpha_i({}_mX_i)$ is large and, in particular, is larger than $r_i({}_mX_i)$, but its significance manifests itself only if $\alpha_i(x)$ is rather small. In principle, this can be accomplished with a successful choice of $\psi_1(x), \dots, \psi_K(x)$ and sufficiently large M. The importance of Proposition 3.3 consists in the fact that the error $r_i({}_mX_i)$, which is the most difficult to control, does not prevent $\widehat{C}_i({}_mX_i)$ from being a lower continuation value.

If α_i are not sufficiently small then the gap between simulated lower and upper bounds is usually large. However, the possibility that the true price lies significantly outside the constructed bounds cannot be completely ruled out. In this case it is hardly possible to judge about α_i on the basis of the gap between lower and upper bounds. Nevertheless, we believe that in most cases the tightness of the bounds do imply a successful estimation of the price. Note, that another regression-based approach suggested in [6] leads always to a "true" upper bound, it does not matter how bad was the choice of basis functions.

In fact, the success of any regression-based approach clearly depends on the choice of basis functions. This is a rather complicated problem, both in practice and in theory. Polynomials (sometimes damped by functions vanishing at infinity) are popular choices (see, e.g., [23] and [28]). Through Taylor expansion, any sufficiently smooth function can be approximated by polynomials. However, the number of monomials of a given degree grows polynomially in the number of variables, so without further assumptions about the structure of the value function the number of basis functions required could grow quickly with the dimension of the underlying state vector. This is why some authors proposed to use special basis functions tailored to the particular problem (see, e.g., [1] or Section 5, where values of European options are employed).

4 Formulas for global lower and upper bounds

Aiming to estimate the price of the American option at a fixed position (t_0, x_0), we simulate the independent trajectories ${}_mX_i$, $i = 1, ..., \mathcal{I}$, $m = 1, ..., M$, of the process X_i, starting at the instant $t = t_0$ from $x_0 : X_0 = x_0$.

To construct the global lower bound, we use the formula (3.12). Indeed, (3.12) gives the following estimate

$$\widehat{u}_0(X_0) = \max\left\{ f_0(X_0), \frac{B_1}{M} \sum_{m=1}^{M} \frac{f_{\widehat{\tau}_m}(X_{\widehat{\tau}_m}^{t_1,\ {}_mX_1})}{B_{\widehat{\tau}_m}} \right\},\ \widehat{\tau}_m = \widehat{\tau}^{t_1,\ {}_mX_1}. \tag{4.1}$$

We note that (4.1) is always a lower bound for $u_0(X_0)$ even if $\widehat{\tau}_m$ is not equal to the optimal stopping time $\tau^{t_1,\ {}_mX_1}$. This estimate weakly depends on the future of the set of trajectories $(t_i, {}_m X_i)$ (see Subsection 3.4). To construct an independent estimate $\breve{u}_0(X_0)$,

we use (3.21) and the estimates ${}^{\breve{\tau}^{t_i,{}_m\breve{X}_i}}$ of the optimal stopping times $\tau^{t_i,{}_m\breve{X}_i}$ introduced in Subsection 3.4:

$$\breve{u}_0(X_0) = \max\left\{ f_0(X_0), \frac{B_1}{\breve{M}} \sum_{m=1}^{\breve{M}} \frac{f_{\breve{\tau}_m}(X_{\breve{\tau}_m}^{t_1,\ {}_m\breve{X}_1})}{B_{\breve{\tau}_m}} \right\},\ \breve{\tau}_m = \breve{\tau}^{t_1,\ {}_m\breve{X}_1}. \tag{4.2}$$

To construct a global upper bound, we use lower continuation values from Section 3. Let $\widehat{C}_i({}_mX_i)$ be a lower continuation value. Then

$$\widehat{\gamma}_i({}_mX_i) = [f_i({}_mX_i) - \widehat{C}_i({}_mX_i)]^+ \tag{4.3}$$

is an upper consumption value and the corresponding global upper bound is given by the formula

$$\widehat{V}_0(x_0) = \frac{1}{M}\sum_{m=1}^{M} \frac{f_{\mathcal{I}}({}_mX_{\mathcal{I}})}{B_{\mathcal{I}}} + \frac{1}{M}\sum_{i=0}^{\mathcal{I}-1}\sum_{m=1}^{M} \frac{\widehat{\gamma}_i({}_mX_i)}{B_i}. \tag{4.4}$$

The independent of future counterpart of (4.4) is the formula

$$\breve{V}_0(x_0) = \frac{1}{\breve{M}}\sum_{m=1}^{\breve{M}} \frac{f_{\mathcal{I}}({}_m\breve{X}_{\mathcal{I}})}{B_{\mathcal{I}}} + \frac{1}{\breve{M}}\sum_{i=0}^{\mathcal{I}-1}\sum_{m=1}^{\breve{M}} \frac{\breve{\gamma}_i({}_m\breve{X}_i)}{B_i},\ \text{with } \breve{\gamma}_i({}_m\breve{X}_i) = [f_i({}_m\breve{X}_i) - \breve{C}_i({}_m\breve{X}_i)]^+. \tag{4.5}$$

The estimates (4.1) and (4.4) are of lower computational cost than (4.2) and (4.5). Due to only weak dependence on the future, (4.1) gives the same results in simulation as (4.2), correspondingly (4.4) gives the same results as (4.5) (see Section 5). That is why we prefer the estimates (4.1) and (4.4) in practice.

Let us recall that the true consumption at (t_i, x) is equal to

$$\gamma_i(x) = [f_i(x) - C_i(x)]^+ \tag{4.6}$$

and $\widehat{\gamma}_i({}_mX_i) \geq \gamma_i({}_mX_i)$ within the accuracy of $\widehat{C}_i({}_mX_i)$ (the accuracy depends on N and M in the case of the local Monte Carlo approach and on K and M in the case of the linear regression approach). If

$$f_i({}_mX_i) < \widehat{C}_i({}_mX_i), \tag{4.7}$$

then $(t_i,\ {}_mX_i) \in \mathcal{C}$ (see (2.2)) and we move one step ahead along the trajectory to the next position $(t_{i+1},\ {}_mX_{i+1})$. Otherwise, if

$$f_i({}_mX_i) \geq \widehat{C}_i({}_mX_i), \tag{4.8}$$

then we cannot say definitely whether the position $(t_i,\ {}_mX_i)$ belongs to $\mathcal{C}$ or $\mathcal{E}$. In spite of this, we make one step ahead in this case as well. In the case (4.7) we have $\widehat{\gamma}_i({}_mX_i) = \gamma_i({}_mX_i) = 0$ and there is no error. If (4.8) holds and besides $\widehat{C}_i({}_mX_i) < C_i({}_mX_i)$, an error can occur. If $\widehat{\gamma}_i({}_mX_i)$ is small, the error is small as well. But if $\widehat{\gamma}_i({}_mX_i)$ is large, then it is, in general, impossible to estimate this error. Note that for the construction of

an estimate $\widehat{V}_0(x_0)$ of the upper bound $V_0(x_0)$ one can use different local lower bounds depending on the position. This opens various opportunities for adaptive procedures (see [4]). For instance, if $\widehat{\gamma}_i({}_mX_i)$ is large, then it is reasonable to use a more powerful local instrument at the position $(t_i,\ {}_mX_i)$.

Remark 4.1. In reality (see (3.20)), the global upper bound is equal to $\widehat{V}_0(x_0) + \Delta$, where $\Delta \to 0$ when $M, N \to \infty$. Therefore we have $\widehat{u}_0(X_0) \le u_0(X_0) \le \widehat{V}_0(x_0) + \Delta$, i.e., the accuracy is estimated by the difference $\widehat{V}_0(x_0) + \Delta - \widehat{u}_0(X_0)$ (not by $\widehat{V}_0(x_0) - \widehat{u}_0(X_0)$). In practice, it may happen that $\widehat{V}_0(x_0) \le \widehat{u}_0(X_0)$. Clearly, in this case the accuracy is bounded by Δ.

5 Simulations

5.1 Bermudan max-calls on d assets

This is a benchmark example studied in [10], [17] and [25] among others. Specifically, the model with d identically distributed assets is considered, where each underlying has dividend yield δ. The risk-neutral dynamic of assets is given by

$$\frac{dX_t^k}{X_t^k} = (r-\delta)dt + \sigma dW_t^k, \quad k = 1, ..., d,$$

where W_t^k, $k = 1, ..., d$, are independent one-dimensional Brownian motions and r, δ, σ are constants. At any time $t \in \{t_0, ..., t_{\mathcal{I}}\}$ the holder of the option may exercise it and receive the payoff

$$f(X_t) = (\max(X_t^1, ..., X_t^d) - K)^+.$$

We take $t_i = iT/\mathcal{I}$, $i = 0, ..., \mathcal{I}$, with $T = 3$, $\mathcal{I} = 9$ and apply the local Monte Carlo method with kernel interpolation scheme as described in Section 3.3. The number of outer Monte Carlo simulations is $M = 10000$ and the number of inner Monte Carlo simulations is $N = 100$. The results are presented in Table 1 depending on x_0 with $X_0 = (X_0^1, \dots, X_0^d)^T$, $X_0^1 = ... = X_0^d = x_0$. The dual upper bounds presented in the third column are computed by the primal-dual algorithm (see [1]), hence by the nested Monte Carlo, with 10000 outer and 100 inner simulations. An initial lower approximation is constructed by the Longstaff-Schwartz method, where all monomials (in X_t) up to order 2 plus the immediate payoff are used in the regression basis (see, [15, p. 476] for comparison). The true values in the last column are quoted from [15] as well. As we can see in this example, the dual method performs slightly better than the local Monte Carlo approach. However, the need for an initial approximation at each point makes the dual approach more expensive. Note that the values of upper bound lie outside 95% confidence interval around the true value. This can be cured by increasing the number of inner simulations N. Figure 5.1, showing bounds $\widehat{u}_0(X_0)$ and $\widehat{V}_0(X_0)$ as functions of N ($d = 2$, $x_0 = 90$), indicates that the lower approximation converges much faster than the upper approximation as N increases.

Table 1: Bounds (with 95% confidence intervals) for the Bermudan max-call with parameters $K = 100$, $r = 0.05$, $\sigma = 0.2$, $\delta = 0.1$ and different d and x_0.

d	x_0	Lower Bound $\widehat{u}_0(X_0)$	Upper Bound $\widehat{V}_0(X_0)$	Dual Upper Bound	True Value
2	90	7.965±0.239	8.417±0.082	8.2311	8.08
	100	13.644±0.300	14.493±0.113	14.182	13.90
	110	20.875±0.370	22.014±0.165	21.681	21.34
5	90	16.795±0.315	19.012±0.153	17.163	16.71
	100	26.265±0.379	29.340±0.183	27.216	26.21
	110	36.790±0.437	40.630±0.208	38.577	36.84

Let us finally note that the quality of the upper bounds becomes worse as d increases. The reason for this is the sparsity of data in $\mathbb{R}^d$ for large d. In such a situation the choice of an interpolating scheme becomes crucial. In the above example we employ the simplest kernel interpolation which is known to be effective for rather small dimensions. For high dimensional spaces, k-nearest neighbors interpolation with an adaptive choice of k (see, [7] for examples and discussions) should be preferred.

5.2 Bermudan swaptions in the Libor market model

Let us consider the Libor market model with respect to a tenor structure $0 = T_0 < T_1 < \ldots < T_{\mathcal{I}}$ in the spot Libor measure P^*. The dynamics of the forward Libor $L_i(t)$, $0 \le t \le T_i$, $i = 1, \ldots, \mathcal{I}-1$, is governed by the SDE (e.g., see [15] and [26])

$$dL_i = \sum_{j=\eta(t)}^{i} \frac{\delta_j L_i L_j \gamma_i^\top \gamma_j}{1+\delta_j L_j}\, dt + L_i\, \gamma_i^\top dW^*, \quad L_i(0) = L_i^0, \quad t \in [0, T_i], \tag{5.1}$$

where $\delta_j = T_{j+1} - T_j$ are day count fractions, $t \mapsto \gamma_i(t) = (\gamma_{i,1}(t), \ldots, \gamma_{i,d}(t))$ are deterministic volatility vector functions defined in $[0, T_i]$ (called factor loadings), and $\eta(t) := \min\{m : T_m > t\}$ denotes the next reset date at time t. In (5.1) $W^*(t)$, $0 \le t \le T_{\mathcal{I}-1}$, is a standard D-dimensional Wiener process under the measure P^* with D, $1 \le D < \mathcal{I}$, being the number of driving factors. The spot Libor measure P^* is induced by the numeraire

$$B^*(t) := \frac{B_{\eta(t)}(t)}{B_1(0)} \prod_{i=1}^{\eta(t)-1} (1 + \delta_i L_i(T_i)), \tag{5.2}$$

where $B_i(t)$, $i = 1, \ldots, \mathcal{I}$, is the value of a zero coupon bond with face value 1 at T_i, i.e., $B_i(T_i) = 1$. At a tenor date T_i, $i = 1, ..., \mathcal{I}-1$, we have (see [15])

$$B_n(T_i) = \prod_{j=i}^{n-1} \frac{1}{1+\delta_j L_j(T_i)}, \quad n = 1, \ldots, \mathcal{I}. \tag{5.3}$$

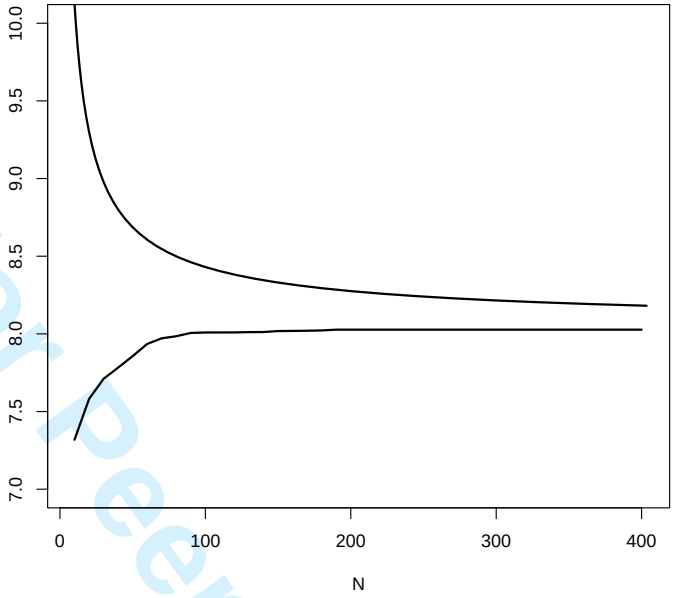

Figure 5.1: Dependence of the local Monte Carlo bounds for the two-dimensional Bermudan max-call on the number of inner Monte Carlo paths N; the number of outer paths M is equal to 10000, $d = 2$, and $x_0 = 90$.

Note that in (5.2) and (5.3) we set by definition $\prod_k^l = 1$ for $k > l$. It is also worth mentioning that $B_n(t)$, $n = 1, \dots, \mathcal{I}-1$, are uniquely defined by Libors on the tenor grid only (we need values of $B^*(t)$ only there as well).

A European swaption with maturity T_i and strike θ gives the right to contract at T_i for paying a fixed coupon θ and receiving floating Libor at the settlement dates $T_{i+1}, \dots, T_{\mathcal{I}}$. The corresponding payoff at maturity T_i is given by

$$f_i(L_i(T_i), \dots, L_{\mathcal{I}-1}(T_i)) := \left(\sum_{j=i}^{\mathcal{I}-1} B_{j+1}(T_i)\delta_j(L_j(T_i) - \theta) \right)^+ .$$

A Bermudan swaption issued at $t = 0$ gives the right to obtain

$$f_i(L_i(T_i), \dots, L_{\mathcal{I}-1}(T_i))$$

at an exercise date $i \in \{s_1, \dots, s_l\} \subset \{1, \dots, \mathcal{I}-1\}$, $s_l = \mathcal{I}-1$, to be decided by the option holder. Its risk-neutral price is given by

$$u_0(L_0(0), \dots, L_{\mathcal{I}-1}(0)) = \sup_{\tau \in \mathcal{T}_S} E\left(\frac{f_\tau(L_\tau(T_\tau), \dots, L_{\mathcal{I}-1}(T_\tau))}{B^*(T_\tau)} \middle| \mathcal{F}_0 \right),$$

where $\mathcal{T}_S$ is the set of stopping times τ taking values in $\{s_1, ..., s_l\}$. The risk-neutral price at a future position $(T_i, L_i(T_i), \dots, L_{\mathcal{I}-1}(T_i))$ provided that the option has not be

exercised before is given by

$$u_i(L_i(T_i),\dots,L_{\mathcal{I}-1}(T_i)) = \sup_{\tau\in\mathcal{T}_S\bigcap[i,\mathcal{I}-1]} B^*(T_i)E\left(\frac{f_\tau(L_\tau(T_\tau),\dots,L_{\mathcal{I}-1}(T_\tau))}{B^*(T_\tau)}\middle|\mathcal{F}_i\right). \tag{5.4}$$

Note that the right-hand side of (5.4) depends only on $(L_i(T_i),\dots,L_{\mathcal{I}-1}(T_i))$ due to the Markov property of the underlying process and the identity

$$\frac{B^*(T_i)}{B^*(T_\tau)} = \frac{1}{\prod_{k=i}^{\tau-1}(1+\delta_k L_k(T_k))}, \quad \tau > i.$$

Hence, the continuation value

$$C_i(L_i(T_i),\dots,L_{\mathcal{I}-1}(T_i)) = B^*(T_i)E\left(\frac{u_{i+1}(L_{i+1}(T_{i+1}),\dots,L_{\mathcal{I}-1}(T_{i+1}))}{B^*(T_{i+1})}\middle|\mathcal{F}_i\right) \tag{5.5}$$

is a function of state vector at time T_i as well. Thus, the results of Sections 3 and 4 remain valid for the considered model though the numeraire is not deterministic.

In our simulation study we use the Libor volatility structure

$$\gamma_i(t) = c_i g(T_i - t)e_i, \tag{5.6}$$

where $g(s) = g_\infty + (1 - g_\infty + as)e^{-bs}$, e_i are D-dimensional unit vectors, decomposing an input correlation matrix of the rank D, and $g_\infty \geq 0$, $a \geq 0$, $b \geq 0$, $c_i > 0$ are constants (see [26]). To generate Libor models with different numbers of factors D, we take as a basis a correlation structure of the form

$$\rho_{ij} = \exp(-\phi|i-j|), \quad i,j = 1,\dots,\mathcal{I}-1,$$

which has full rank for $\phi > 0$. Then for a particular choice of D a rank-D correlation matrix $\rho^{(D)}$ with decomposition $\rho^{(D)}_{ij} = e_i^\top e_j$, $1 \leq i,j < \mathcal{I}$, is obtained from ρ by principal component analysis. As the model parameters, we take a flat 10% initial Libor curve (i.e. $L_i^0 = 0.1$ for $i = 0,1,\dots,\mathcal{I}-1$) over a 40 period quarterly tenor structure and

$$\mathcal{I} = 41,\ \delta_i = 0.25,\ c_i = 0.2,\ a = 1.5,\ b = 3.5,\ g_\infty = 0.5,\ \phi = 0.0413.$$

We consider Bermudan swaptions with yearly exercise opportunities, hence (δ_i are equal to a quarter year) $s_i = 4i$, $\quad i = 1,\dots,10$. For a "practically exact" numerical integration of the SDE, we use the log-Euler scheme with $\Delta t = \delta/5 = 0.05$.

Now we apply the regression method described in Section 3.5. At each exercise date T_{s_i}, $i = 1,\dots,l-1$, the set of basis functions includes the value of the European swaption

$$\mathcal{S}_i(L_{s_i}(T_{s_i}),\dots,L_{\mathcal{I}-1}(T_{s_i})) = B^*(T_{s_i})E\left(\frac{f_{s_{i+1}}(L_{s_{i+1}}(T_{s_{i+1}}),\dots,L_{\mathcal{I}-1}(T_{s_{i+1}}))}{B^*(T_{s_{i+1}})}\middle|\mathcal{F}_{s_i}\right),$$

which we can be exercised at the next exercise date $T_{s_{i+1}}$, and the immediate payoff f_{s_i} with its second-order powers. Although closed form expressions for European swaptions

do not exist in a Libor market model, there exist very accurate (typically, better than 0.3% relative error) formulas (see [26]), which we use for the computation of $\mathcal{S}_i$.
The resulting lower bound $\widehat{u}_0$ and upper bound $\widehat{V}_0$ are given in Table 2 for different numbers of factors D and different coupons θ. The true values (computed with less than 1% relative error) are quoted from [21].

D	θ	$\widehat{u}_0$	$\widehat{V}_0$	True Value
1	0.08	1108.1±1.5	1110.5±2.4	1109.2
	0.10	381.7±1.2	384.7±1.6	382.1
	0.12	121.2±0.7	123.1±0.8	121.3
10	0.08	1096.3±1.3	1096.6±2.0	1096.5
	0.10	344.3±1.0	346.7±1.3	344.7
	0.12	101.7±0.6	104.9±0.7	101.3
40	0.08	1094.8±1.2	1096.1±2.0	1096.1
	0.10	338.2±1.0	341.2±1.3	339.3
	0.12	96.4±0.5	100.0±0.6	97.2

Table 2: Prices of Bermudan swaptions $\times 10^4$

References

[1] L. Andersen and M. Broadie (2001). A primal-dual simulation algorithm for pricing multidimensional American options. Management Sciences, **50**, No. 9, 1222-1234.

[2] V. Bally, G. Pagès, J. Printems (2005). A quantization tree method for pricing and hedging multidimensional American options. Mathematical Finance, **15**, No. 1, 119-168.

[3] D. Belomestny, G.N. Milstein (2006). Monte Carlo evaluation of American options using consumption processes. International Journal of Theoretical and Applied Finance, **9**, No. 4, 1-27.

[4] D. Belomestny, G.N. Milstein (2005). Adaptive simulation algorithms for pricing American and Bermudan options by local analysis of financial market. WIAS-Preprint No. 1022, Berlin.

[5] D. Belomestny, G.N. Milstein, V. Spokoiny (2006). Regression methods in pricing American and Bermudan options using consumption processes. WIAS-Preprint No. 1145, Berlin.

[6] D. Belomestny, Ch. Bender, J. Schoenmakers (2007). True upper bounds for Bermudan products via non-nested Monte Carlo, to appear in Mathematical Finance.

[7] D. Belomestny and V. Spokoiny (2004). Local-likelihood modeling via stage-wise aggregation. Annals of Statistics, **35**(5), 2287–2311.

[8] P. Boyle, M. Broadie, P. Glasserman (1997). Monte Carlo methods for security pricing. Journal of Economic Dynamics and Control, **21**, 1267-1321.

[9] M. Broadie and J. Detemple (1996), American option valuation: New bounds, approximations and a comparison of existing methods. The Review of Financial Studies, **9**, 1211–1250.

[10] M. Broadie, P. Glasserman (1997). Pricing American-style securities using simulation. J. of Economic Dynamics and Control, **21**, 1323-1352.

[11] J. Carriere (1996). Valuation of early-exercise price of options using simulations and nonparametric regression. Insuarance: Mathematics and Economics, **19**, 19-30.

[12] E. Clément, D. Lamberton, P. Protter (2002). An analysis of a least squares regression algorithm for American option pricing. Finance and Stochastics, **6**, 449-471.

[13] R.A. DeVore (1998). Nonlinear approximation. Acta Numer., **7**, 51-150.

[14] D. Egloff (2005). Monte Carlo algorithms for optimal stopping and statistical learning. Ann. Appl. Probab. **15**, no. 2, 1396-1432.

[15] P. Glasserman (2004). Monte Carlo Methods in Financial Engineering. Springer.

[16] L. Györfi, M. Kohler, A. Krzyżak, H. Walk (2002). A Distribution-Free Theory of Nonparametric Regression. Springer.

[17] M. Haugh, L. Kogan (2004). Pricing American options: a duality approach. Opeations Research, **52**, No. 2, 258–270.

[18] F. Jamshidian (2003). Minimax optimality of Bermudan and American claims and their Monte Carlo upper bound approximation. Stochastics **79**, no. 1-2, 27–60.

[19] V. Kargin (2005). Lattice option pricing by multidimensional interpolation. Math. Finance, **15**, No. 4, 635–647.

[20] A. Kolodko, J. Schoenmakers (2004). Upper bounds for Bermudan style derivatives. Monte Carlo Methods and Appl., **10**(3-4), 331-343.

[21] A. Kolodko, J. Schoenmakers (2006). Iterative construction of the optimal Bermudan stopping time. Finance and Stochastics, **10**, No. 1, 27-49.

[22] D. Lamberton, B. Lapeyre (1996). Introduction to Stochastic Calculus Applied to Finance. Chapman & Hall.

[23] F.A. Longstaff, E.S. Schwartz (2001). Valuing American options by simulation: a simple least-squares approach. Review of Financial Studies, **14**, 113-147.

[24] G.N. Milstein, M.V. Tretyakov (2005). Numerical Analysis of Monte Carlo evaluation of Greeks by finite differences. Journal of Computational Finance, **8**, No. 3, 1-33.

[25] L.C.G. Rogers (2001). Monte Carlo valuation of American options. Mathematical Finance, **12**, 271-286.

[26] J. Schoenmakers (2005). Robust Libor Modelling and Pricing of Derivative Products. Chapman & Hall/CRC.

[27] A.N. Shiryaev (1999). Essentials of Stochastic Finance: Facts, Models, Theory. World Scientific.

[28] J. Tsitsiklis, B. Van Roy (1999). Regression methods for pricing complex American style options. IEEE Trans. Neural. Net., **12**, 694-703.

[29] H. Wendland (2005). Scattered Data Approximation. Cambridge: Cambridge University Press.